# Fisher Error Matrix

- The Fisher matrix is defined by the expectation value:

$$\boxed{\mathbf{F}_{ij} = \left\langle -\frac{\partial^2 \ln L(x,\Theta)}{\partial\Theta_i \partial\Theta_j} \right\rangle = \left\langle \frac{\partial \ln L(x,\Theta)}{\partial\Theta_i} \frac{\partial \ln L(x,\Theta)}{\partial\Theta_j} \right\rangle}$$

  – where $L(x,\Theta) = \prod_{k=1}^{K} f(x_k,\Theta)$ is the combined probability distribution (or likelihood function); and $f(x_k,\Theta)$ is the probability distribution of the individual measurements $x_k$ that in general also depends on the model parameters $\Theta$.

# Why Fisher?

- It can be shown that:
  - For any unbiased estimator (the set of parameters that won't bias the determination of the true parameters and gives as small error bars as possible)

$$\boxed{\begin{aligned} &\Delta\Theta_i \geq 1\Big/\sqrt{F_{ii}}\text{, where} \\ &\Delta\Theta_i = \sqrt{\left\langle \Theta_i^2 - \langle\Theta_i\rangle^2 \right\rangle} \end{aligned}}$$

**Cramer-Rao inequality**

    - In other words, there is a firm lower limit on the error bars that can be attained.
  - **For a very large dataset, the maximum likelihood estimator (the parameter set that maximizes the likelihood), is the best estimator, and that's the one that makes the Cramer-Rao inequality an equality.**

# The Fisher Information Matrix

- We can hand-wave a proof of the Cramer-Rao inequality
- Let's Taylor expand $\mathcal{L}$ (the likelihood) around the maximum likelihood estimator QML.

$$\boxed{\mathcal{L} = \mathcal{L}(\Theta_{ML}) + \cancelto{0}{\left.\frac{\partial \mathcal{L}}{\partial \Theta_i}\right|_{\Theta=\Theta_{ML}}} \Delta\Theta_i + \frac{1}{2} \left.\frac{\partial^2 \mathcal{L}}{\partial \Theta_i \partial \Theta_j}\right|_{\Theta=\Theta_{ML}} \Delta\Theta_i \Delta\Theta_j + \ldots b}$$

- Because $\mathcal{L}$ = -ln($L$), so $L$ = exp(-$\mathcal{L}$), $\mathcal{L}$ is roughly Gaussian around the ML point. Therefore, we see that the covariance matrix is just:

$$\boxed{\left(T^{-1}\right)_{ij} \equiv \frac{\partial^2 \mathcal{L}}{\partial \Theta_i \partial \Theta_j}}$$

# Recall: Normal Distribution

- The probability density of the normal distribution is given by:

$$N(x; m, s^2) = \frac{1}{\sqrt{2ps^2}} \exp\left(-\frac{(x-m)^2}{2s^2}\right) \qquad (1)$$

  – where the mean is μ, the variance is σ².

- For the multi-variable case:

$$N(\mathbf{x}; \boldsymbol{\mu}, \mathbf{C}) = \frac{1}{\sqrt{2p}} |\mathbf{C}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \mathbf{C}^{-1} (\mathbf{x}-\boldsymbol{\mu})\right) \qquad (2)$$

  – where m is the mean vector, and **C** is the covariance matrix
  – If covariance matrix is diagonal, the multivariate normal distribution reduces to a product of independent univariate normal distributions

# The Gaussian Case

- If the probability distribution *L* is Gaussian (given by Eqn. 2 on the previous slide), then $\mathcal{L}$ is given by:

$$\boxed{2\mathcal{L} = \ln \det C + (\mathbf{x}-\boldsymbol{\mu})^T \mathbf{C}^{-1} (\mathbf{x}-\boldsymbol{\mu}) + const}$$

- From here, with lots of matrix manipulations and using the definition of the Fisher (see the first slide), one can show that

$$\boxed{\mathbf{F}_{ij} = \frac{1}{2} \mathrm{Tr}\left[\mathbf{C}^{-1}\mathbf{C}_{,i}\mathbf{C}^{-1}\mathbf{C}_{,j}\right] + \boldsymbol{\mu}^T_{,i}\mathbf{C}^{-1}\boldsymbol{\mu}_{,j}}$$

  - Where m is the mean vector of the data, **C** is the covariance matrix, and commas denote derivatives: $\mathbf{C}_{,i} \equiv \frac{\partial}{\partial \Theta_i}\mathbf{C}$

# Application 1: Cosmic Complementarity

- The "Cosmic Complementarity" paper by Tegmark *et al.* gives a very nice example of using the Fisher information matrix for understanding the complementarity of the cosmological tests
  - They assume the Gaussian case, making it much simpler
- Consider a number of noisy measurements $\mathbf{x}_\mathbf{n}$ (e.g., for SN Ia, those are the observed magnitudes of n objects). The parameters $\theta$ that we are trying to get at are $\Omega_\Lambda$ and $\Omega_M$. Then, the 2x2 Fisher matrix is given by:

$$\boxed{\mathbf{F}_{ij} = a^2 \sum_{n=1}^{N} \frac{1}{s^2} \frac{\partial \ln d(z_n)}{\partial \Theta_i} \frac{\partial \ln d(z_n)}{\partial \Theta_j}}$$

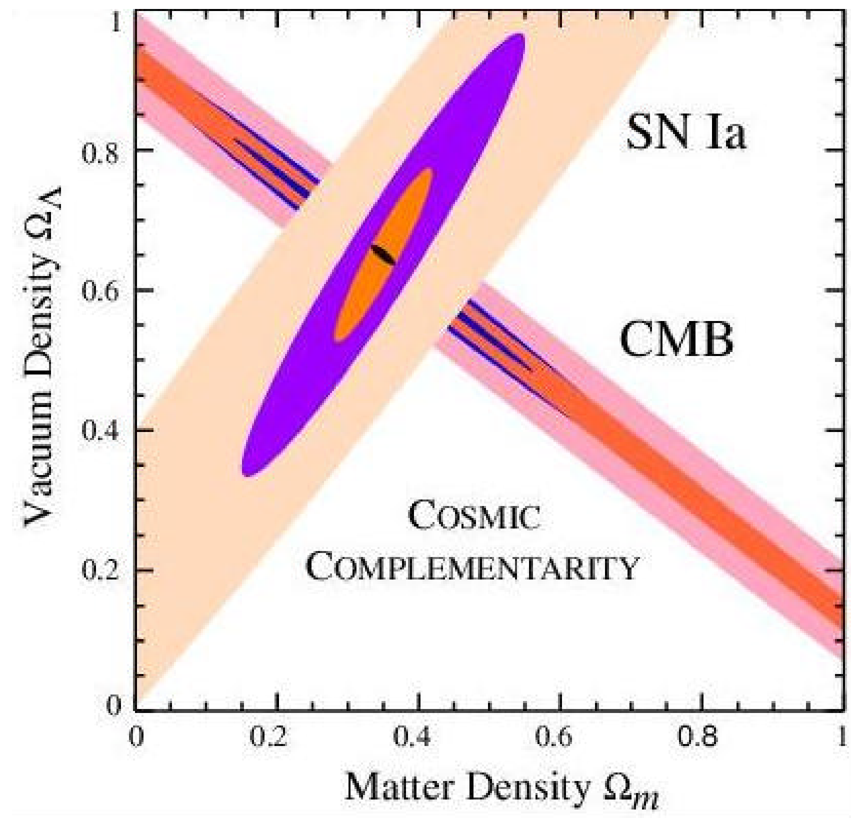
  - where $a$ = const, $d$ = luminosity distance, and $n$ goes over $N$ redshifts.
- Explicitly evaluating this Fisher, we get the errors on the two variables, which we then plot as error ellipses. We can do the same for the CMB, and it turns out that the error ellipses are almost orthogonal to those of SN.

# Cosmic Complementarity

- The contribution to the Fisher from each redshift reflects the quality of the dataset & incorporates the effects of cosmology (Eqn. 10 in the paper)
- For the CMB, although the paper doesn't say anything explicitly, it can be shown that the Fisher is:

$$\mathbf{F}_{ll'} = \frac{2l+1}{2} d_{ll'} e^{-2l^2 s^2} \left[ C_l e^{-l^2 s^2} + w^{-1} \right]^{-2}$$

  - Where $C_l$ is the component of the angular power spectrum for multipole number $l$, $\sigma$ is the beam size, and $w$ is the weight that includes pixel size and noise.



# Application 2: the ETC

- In Gary Berstein's paper on the advanced ETC, he points out that the Fisher information matrix for point source photometry is given by:

$$\boxed{\mathbf{F}_{ij} = \sum_{\text{pixels}} \frac{\left|\partial(fP')/\partial p_i\right|\left|\partial(fP')/\partial p_j\right|}{\mathrm{Var}(I(\mathbf{x}_i))}}$$

  - Here, $\boldsymbol{p} = \{f, \; x_0, \, y_0\}$ are the parameters we want to know (flux and position), $I$ is the measured flux at position $\boldsymbol{x}_i$, and $P'$ is the (effective) PSF.
  - In particular, if we know the centroid $\{x_0, \, y_0\}$, then the error on the flux is just the inverse of the (0,0) component of the Fisher: $(F_{ff})^{-1}$.

# The Advanced ETC

- Using the Fisher matrix method, Gary derives a number of interesting conclusions:
  - If pixels are larger than the Nyquist size (0.5 $\lambda$/D), the optimal amount of dithering should be about half the Nyquist density (the interlace factor N should be 2-3).
  - The pixel scale should be >=1 for survey-oriented projects. It also helps to have larger pixels for point-source astrometry, and for measurements of galaxy ellipticities.
  - For faint and far away sources, space begins to be clearly advantageous over ground. In the Z-band, the space advantage is 7-10 times; in the NIR the advantage is huge.